

# ESTIMATION PAR ONDELETTES DANS DES MODÈLES FONCTIONNELS GÉNÉRALISÉS

Irène Gannaz

*Université de Lyon, CNRS UMR 5208, INSA de Lyon, Institut Camille Jordan,  
20, avenue Albert Einstein, 69 621 Villeurbanne cedex, France.  
E-mail: irene.gannaz@insa-lyon.fr*

**Résumé.** L'estimation par ondelettes de signaux en présence de bruit gaussien a été largement développée ces dernières années. Le but de ce travail est d'étendre les résultats à des contextes faisant appel à des distributions plus générales telles que Poisson, Binomiale ou Gamma... Nous considérons une approche par log-vraisemblance pénalisée, où la pénalité s'exprime à l'aide des coefficients d'ondelettes. Nous nous intéressons par ailleurs au cas où un terme linéaire est estimé simultanément. Nous montrons l'optimalité asymptotique de la procédure d'estimation et nous proposons un algorithme simple de mise en oeuvre.

**Mots-clés.** Modèles semiparamétriques, régression généralisée, ondelettes, seuillage, log-vraisemblance pénalisée.

**Abstract.** We consider a generalized functional regression model with an unknown nonparametric function. We present a maximum penalized likelihood procedure introducing penalty based wavelet estimators. A semiparametric extension, with the presence of a linear part in the model, is also considered. Asymptotic rates of the estimates of both the parametric and the nonparametric part of the model are given and quasi-minimax optimality is obtained under usual assumptions in literature. We establish in particular that the  $\ell^1$ -penalty leads to an adaptive estimation. An algorithm is also proposed for implementation and simulations are used to illustrate the results.

**Keywords.** Semiparametric models, generalized regression, penalized loglikelihood estimation, wavelet thresholding.

## Introduction

Nous nous intéressons ici à un modèle de régression généralisée, où l'on souhaite prédire l'influence de variables explicatives sur une variable d'intérêt à l'aide d'un lien de type fonctionnel. Nous autorisons par ailleurs la présence d'une partie linéaire dans le modèle.

Dans le contexte gaussien, l'estimation par ondelettes a permis le développement d'estimateurs dits adaptatifs par rapport à la régularité de la fonction, c'est-à-dire d'estimateurs dont les

paramètres sont pré-déterminés et ne nécessitent pas d'étape de validation croisée afin de choisir leur valeur adéquate.

Le but du travail exposé est de développer une estimation similaire lorsque les observations ne sont plus issues d'une loi gaussienne mais d'une loi de la famille exponentielle. Ainsi, pour des observations suivant des lois de Poisson, des lois binomiales, Gamma, *etc*, nous proposons une méthode d'estimation par ondelettes adaptative et nous montrons sa quasi-optimalité asymptotique au sens minimax. En outre, nous verrons que cette procédure d'estimation présente le grand intérêt d'une mise en œuvre algorithmique aisée.

## 1 Modèle et procédure d'estimation

Nous nous intéressons dans ce travail à un modèle de régression fonctionnelle généralisée. Le but est de prédire une variable  $Y$  en fonction des variables explicatives  $\mathbf{z}$ , en relâchant l'hypothèse usuelle de normalité de la réponse. Plus précisément, nous supposons que  $y$  est la réalisation d'une variable unidimensionnelle de loi exponentielle dont la densité est de la forme :

$$\exp\left(\frac{y\eta(\mathbf{z}) - b(\eta(\mathbf{z}))}{\phi} + c(y, \phi)\right). \quad (1)$$

Dans cette expression,  $b(\cdot)$  et  $c(\cdot)$  sont des fonctions données qui caractérisent la forme de la distribution. Le paramètre  $\phi$  est un paramètre de dispersion que nous supposerons connu. Quant à la fonction  $\eta(\cdot)$ , elle contient l'information relative à l'influence des variables explicatives. Etant donné un échantillon indépendant de  $n$  observations issues de ce modèle, notre objectif est d'estimer cette fonction  $\eta(\cdot)$ .

Nous considérons  $\eta(\cdot)$  sous une forme fonctionnelle, mais nous autorisons la présence de covariables d'influence linéaire. Nous obtenons ainsi le modèle semi-paramétrique :

$$\eta(\mathbf{X}, t) = \mathbf{X}^T \boldsymbol{\beta} + f(t), \quad (2)$$

où  $\mathbf{z} = (\mathbf{X}, t) \in \mathbb{R}^p \times \mathbb{R}$  avec  $\boldsymbol{\beta}$  un vecteur de dimension  $p$  et  $f(\cdot)$  une fonction à valeur réelle. Etant données des observations  $(Y_i, \mathbf{X}_i, t_i)_{i=1, \dots, n}$ , le but est de développer une procédure d'estimation du vecteur  $\boldsymbol{\beta}$  et de la fonction  $f(\cdot)$ .

Ce modèle offre un large éventail d'applications possibles. En effet, de nombreuses distributions sont envisagées : Gaussienne, Poisson, Binomiale, Gamma, *etc*. Pour plus de références sur la régression généralisée et ses applications nous renvoyons au livre de Fahrmeir et Tutz (1994). De nombreux exemples d'applications dans le cas fonctionnel peuvent par ailleurs être trouvés dans l'article de Schwartz (1994).

Afin d'estimer simultanément le vecteur  $\boldsymbol{\beta}$  et la fonction  $f(\cdot)$  à partir des observations  $(Y_i, \mathbf{X}_i, t_i)_{i=1, \dots, n}$ , nous optons pour une approche par logvraisemblance pénalisée. Soit  $\ell$  la

logvraisemblance du modèle :  $\ell(y, \eta) = y\eta - b(\eta)$ . Les estimateurs  $\hat{f}_n$  et  $\hat{\beta}_n$  sont solutions du problème d'optimisation :

$$\begin{aligned} (\hat{f}_n, \hat{\beta}_n) &= \underset{\{f, \|f\|_\infty \leq C_\infty\}, \beta \in \mathbb{R}^p}{\operatorname{argmax}} K_n(f, \beta) \\ \text{avec } K_n(f, \beta) &= \sum_{i=1}^n \ell(y_i, \mathbf{X}_i^T \beta + f(t_i)) - \lambda \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} |\theta_{j,k}^W|. \end{aligned}$$

Les coefficients  $\theta_{j,k}^W$  sont les coefficients de la fonction  $f(\cdot)$  obtenus par une décomposition en ondelettes ;  $\theta_{j,k}^W = \Psi(f(t_1) \dots f(t_n))^T$  où  $\Psi$  est la transformée en ondelettes discrète.

Dans le cadre gaussien, nous savons que la pénalité  $\ell^1$  ci-dessus permet d'obtenir l'estimateur non linéaire par ondelettes usuel avec un seuillage doux de seuil  $\lambda$  sur les coefficients. Nous montrons dans quelle mesure ceci est aussi applicable dans le cadre d'une famille exponentielle : cette pénalité correspond en fait à un seuillage doux itératif, comme ceci sera vu en section 3.

Un point intéressant est que cette pénalité, pour une distribution gaussienne comme généralisée, mène à une estimation adaptative. En effet, nous verrons que le choix du paramètre  $\lambda$  permettant d'obtenir une vitesse optimale au sens minimax ne nécessite pas de connaissance quant à la régularité de la fonction  $f(\cdot)$ .

## 2 Comportement asymptotique

Soit  $B$  la matrice diagonal dont le  $i^{\text{ème}}$  terme vaut  $\ddot{b}(\eta(\mathbf{X}_i \beta + f(t_i)))$ . Introduisons  $H = X(X^T B X)^{-1} X^T$  la matrice de projection sur l'espace engendré par les colonnes de  $\mathbf{X}$  pour le produit scalaire associé à la distribution (1). La matrice  $H$  est de rang et de trace  $p$ . Notons  $h_i = \mathbf{X}_i(X^T B X)^{-1} \mathbf{X}_i^T$ . Alors  $\sum h_i = p$  et nous supposons aussi que  $h = \max_{i=1 \dots n} h_i \rightarrow 0$ .

Nous émettons ensuite quelques hypothèses, nécessaires à l'étude du comportement asymptotique.

**(H1)**  $\frac{1}{n} \sum_{i=1}^n \ddot{b}(\eta_{0,i}) X_i X_i^T$  converge vers une matrice positive quand  $n$  tend vers l'infini et  $\frac{1}{n} X^T F$  tend vers avec  $\mathbf{f} = (f(t_1), \dots, f(t_n))^T$ .

**(H2)**  $\sup_{\|\eta - \eta_0\|_n \leq 2(C_\infty + \|f\|_\infty)} \sup_i \ddot{b}(\eta_i) \leq \ddot{b}_\infty < \infty$ .

Il existe  $K, \sigma_0^2 > 0$  tels que

$$\max_{i=1, \dots, n} K^2 \left( \mathbb{E}[\exp \left( \left| \dot{\ell}(Y_i, \mathbf{X}_i^T \beta + f(t_i)) \right| / K^2 \right)] - 1 \right) \leq \sigma_0^2.$$

La première hypothèse sert à assurer l'identifiabilité du modèle. Quant à la seconde, elle signifie que les lois considérées ont des queues sous-gaussiennes.

Nous voulons enfin contrôler la corrélation entre la partie linéaire et la partie fonctionnelle. De manière similaire aux travaux de Rice (1984) et Speckman (1986), nous décrivons les vecteurs de la matrice de régression  $X$  en faisant intervenir une décomposition de type fonctionnel.

**(H3)**  $\forall j = 1, \dots, p, i = 1, \dots, n, X_{i,j} = g_j(t_i) + \xi_{i,j}$ , avec  $g_j$  fonctions polynomiales de degrés inférieurs ou égaux au nombre de moments nuls des ondelettes. Pour tout  $j$  les variables  $\xi_{i,j}$  sont supposées indépendantes issues de lois subgaussiennes.

Nous sommes maintenant en mesure d'énoncer le comportement asymptotique des estimateurs.

**Théorème.** *Supposons les hypothèses (H1) à (H3) vérifiées. Supposons aussi que  $f(\cdot)$  appartienne à une boule de Besov  $\mathcal{B}_{\pi,r}^s(C)$  avec  $s+1/\pi-1/2 > 0$  et  $1/2 < s < \min(R, N)$ , où l'ondelette  $\psi$  est  $R$ -régulière et admet  $N$  moments nuls. Pour  $\lambda = c\sqrt{\log(n)}$ , nous avons*

$$\begin{aligned} \left( \frac{n}{\log(n)} \right)^{s/(1+2s)} \|\tilde{f}_{\beta} - f_0\|_n &= \mathcal{O}_P(1) \\ \sqrt{n} \|\hat{\beta}_n - \beta_0\| &= \mathcal{O}_P(1). \end{aligned}$$

L'optimalité asymptotique au sens minimax est obtenue, et pour l'estimateur de la partie fonctionnelle, et pour l'estimateur du vecteur de régression linéaire. De plus les conditions de corrélation sont similaires à celles obtenues dans le cadre gaussien par Gannaz (2007) et moindres que Mammen et Van Der Geer (1997).

De plus, Rice (1984) et Speckman (1986) s'accordent sur le fait qu'une procédure de validation croisée dans ce contexte semi-paramétrique doit être évitée. Or, nous remarquons que le choix du paramètre de lissage  $\lambda$  ne dépend pas de la régularité de la fonction et que la procédure est donc adaptative. Remarquons toutefois que ce résultat est asymptotique et qu'une application à un échantillon fini nécessitera un choix de la constante  $c$ , qui dépendra du type de distribution considéré.

### 3 Applications sur simulations

Nous appliquons un algorithme de type *backfitting* dont le principe est d'estimer successivement  $f(\cdot)$  puis  $\beta$ . Chacune de ces étapes est résolue à l'aide d'un algorithme de Fisher-scoring, usuel dans les modèles généralisés. Nous décrivons ici la  $k^{\text{ème}}$  étape après application de Fisher-scoring. Rappelons les notations  $\eta(X, t) = X^T \beta + f(t)$  et  $\mu(X, t) = \dot{b}(\eta(X, t))$ . Nous omettons la dépendance en  $(X, t)$  afin de simplifier l'écriture.

1. **Première étape :** Soit  $Y^{(k)}$  le vecteur des pseudo-variables obtenues par descente du gradient et soit  $W^{(k)}$  la matrice (diagonale) de variance de ces pseudo-variables. La fonction  $f^{(k+1)}$  est obtenue par estimation non linéaire par ondelettes à partir des observations

$Y^{(k)}$ , en appliquant un seuillage doux sur les coefficient d'ondelettes. Les seuils sont donnés par  $\lambda \Psi W^{(k)-1} \Psi^T \mathbb{1}_{n \times 1}$  où  $\Psi$  est la transformée en ondelettes et  $\Psi^T$  son inverse.

2. **Deuxième étape :** Soit  $\tilde{Y}^{(k)}$  le vecteur des pseudo-variables obtenues par descente du gradient après avoir remplacé  $f(\cdot)$  par son estimation et soit  $\tilde{W}^{(k)}$  la matrice (diagonale) de variance de ces pseudo-variables. Alors  $\beta^{(k+1)}$  est défini comme le paramètre de régression de  $\tilde{Y}^{(k)}$  par rapport à  $X$  avec les pondérations  $\tilde{W}^{(k)}$ .

Cet algorithme ne nécessite pas d'étape coûteuse ; notamment la seule inversion de matrice nécessaire a lieu sur une matrice diagonale.

Antoniadis, Besbeas et Sapatinas (2001), Antoniadis et Sapatinas (2001) et plus récemment Brown, Cai et Zhou (2010) ont proposé des procédures d'estimation dans de tels modèles (sans présence de partie linéaire). Dans ces articles, une étape préalable consiste à agréger les données en sous-intervalles. Si dans les deux premiers articles cela est peu contraignant, notons que cela demande un nombre très important d'observations pour le troisième. Par comparaison notre algorithme présente une plus grande simplicité.

De plus, l'étude des simulations montre que l'on peut en effet choisir le même paramètre  $\lambda$  pour différentes fonctions. Cette stabilité du paramètre de lissage est un avantage intéressant de cette procédure. Ceci n'est par exemple pas vérifié dans l'algorithme proposé par Sardy, Antoniadis et Tseng (2004) dans un cadre de régression fonctionnelle généralisée, ni par l'estimation de Reynaud-Bouret et Rivoirard (2010) dans un modèle de Poisson fonctionnel.

Enfin, l'étude menée sur simulations montre que la qualité du schéma d'estimation proposé sur des échantillons de taille finie apparaît raisonnable. La reconstitution de la partie fonctionnelle donne également une qualité visuelle correcte.

## Conclusion

Nous proposons ici une estimation par logvraisemblance pénalisée dans un modèle partiellement linéaire généralisé. Avec une pénalité de type  $\ell^1$  sur les coefficients d'ondelettes de la partie fonctionnelle, nous obtenons une procédure adaptative par rapport à la régularité de la fonction. Nous établissons le caractère presque-minimax des estimateurs pour la partie linéaire comme fonctionnelle. Ce résultat est valable pour une large gamme de fonctions, possiblement non régulières, et sous des conditions de corrélation similaires à celle rencontrées dans la littérature dans un contexte gaussien. Par ailleurs, nous présentons l'implémentation des estimateurs et montrons des résultats satisfaisants sur des simulations.

## Bibliographie

- [1] Antoniadis A., Sapatinas T. (2001), Wavelet shrinkage for natural exponential families with quadratic variance functions. *Biometrika*, No. 88, 805–820.
- [2] Antoniadis A., Besbeas P., Sapatinas T. (2001) Wavelet shrinkage for natural exponential families with cubic variance functions. *Sankhyā: The Indian Journal of Statistics*, Vol. 363, No. 3, 309–327.
- [3] Brown L., Cai T., Zhou H. (2010), Nonparametric regression in exponential families. *The Annals of Statistics*, Vol. 38, No. 4, 2005–2046.
- [4] Fahrmeir L., Tutz G. (1994), *Multivariate statistical modelling based on Generalized Linear Models*. Springer.
- [5] Gannaz I. (2007), Robust estimation and wavelet thresholding in partially linear models. *Statistics and Computing*, Vol. 4, No. 17, 293–310.
- [6] Mammen E., Van der Geer S. (1997), Penalized quadi-likelihood estimation in partial linear models. *The Annals of Statistics*, Vol. 25, No. 3, 1014–1035.
- [7] Reynaud-Bouret P., Rivoirard V. (2010), Near optimal thresholding estimation of a Poisson intensity on the real line. *Electronic Journal of Statistics*, Vol. 4, 172–238.
- [8] Rice J. (1986), Convergence rates for partially splined models. *Statistics and Probability Letters*, Vol. 4, 203–208.
- [9] Sardy S., Antoniadis A., Tseng P. (2004), Automatic smoothing with wavelets. *Journal of computational and graphical statistics*, Vol. 13, No. 2, 1–23.
- [10] Shwartz J. (1994), Nonparametric Smoothing in the Analysis of Air Pollution and Respiratory Illness, *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, Vol. 22, No. 4, 471–487.
- [11] Speckman P. (1988), Kernel smoothing in partial linear models. *Journal of Royal Statistical Society*, Vol. 50, No. 3, 413–436.